Due: Tuesday, October 9, 11:10 AM, in class

You may discuss this assignment with whomever you wish, but please prepare and submit work in teams of ***no more than three*** people. While team submissions are encouraged, individually prepared solutions are equally acceptable. Please indicate all team members' names on your team assignment; one submission per team, please. Please prepare solutions in a ***neat, organized fashion***. I prefer typeset presentations (e.g., cut and paste code/output into MS Word with added exposition when appropriate; knitr via EMACS and ESS, knitr or R Markdown via Rstudio, etc.)—probably most appropriate for presenting (fixed width font) code/output, at least—but neatly handwritten presentations may also be appropriate for some problems. Sloppily prepared solutions will not receive full credit.

NOTE: You will be asked to provide a peer assessment for each of your team members—not including yourself!—by assigning each member a score from 0 to 10, 10 being best. Each team member should submit an 8.5 x 11 (holed) notebook sheet with your name, assignment number and the names and scores of each of your team members. Share these assessments only with me, please.

In the Big Bang example of our notes (§6.8), we conducted a few inferences for the regression parameters as well as for the regression function, $\mathrm{E}(Y \mid x)$, and for an unobserved value, $Y \mid x$.

1. Formally write down a null hypothesis and alternative hypothesis (as generally stated in §6.7.2) for testing if the intercept is zero or not (as computed in the code in §6.8). ANS:

$$\begin{aligned} \mathrm{H}_0 : \mathbf{C}\boldsymbol{\beta} &= \mathbf{b}_0 \\ \mathrm{H}_a : \mathbf{C}\boldsymbol{\beta} &\neq \mathbf{b}_0 \qquad \text{where} \\ \mathbf{C} &= \begin{bmatrix} 1 & 0 \end{bmatrix}, \\ \boldsymbol{\beta} &= \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \qquad \text{and} \\ \mathbf{b}_0 &= [0]. \end{aligned}$$

   Formally, we reject the null hypothesis ($\alpha = 0.05$) and conclude that the data do not support the theory that the intercept is zero (F=11.3 (t=3.36), p-value = 0.0028). (As already computed in the code in §6.8.)

2. An interval estimate is often recommended along with test results to help indicate the "importance" of a statistically significant test result.

   (A large data set will often give statistically significant results for an unimportant departure from a null hypothesis if we do not adjust (make smaller) the level of the test ("α") as sample size, n, increases. In other words, a fixed α implies the increasing importance of a Type II error ("β") as sample size increases. See the discussion in [Wak13, §4.2.3 and 4.2.4] about the notion of an *(in)consistent* test procedure and the balancing of Type I and Type II errors by reducing the level of the Type I error probability as the Type II error probability (for some specified minimally important departure presumably) decreases with increasing sample size, n.)

(a) Modify (slightly!) the "by hand $\mathbf{C}\boldsymbol{\beta}$" code given in our notes for the Big Bang example in §6.8 to compute an interval estimate for the intercept, β_0, the distance of a nebula from Earth, traveling from each other at 0 velocity. The simple theory presented in this example says that that distance should be zero, as we discussed in class. ANS: We estimate the intercept to be 0.3992, within (0.1531, 0.6453) with 95% confidence (by default).

```
> bigbang.df<- Sleuth3::case0701
> y<- bigbang.df$Distance
> X<- model.matrix(Distance ~ Velocity, data=bigbang.df)
> (n<-dim(X)[1]); (p<- dim(X)[2]);

[1] 24
[1] 2

> (betahat<- (XtXinv<- solve( XtX<-t(X)%*%X ))%*%t(X)%*%y)

                   [,1]
(Intercept) 0.399170440
Velocity    0.001372408

> ehat<- (y - X%*%betahat)
> (sig2hat<- as.vector(t(ehat) %*% ehat / (n - p)))

[1] 0.1645359

> sqrt(sig2hat)

[1] 0.4056302

> varBhat<- sig2hat * XtXinv
> (seBhat<- sqrt(diag(varBhat)))

 (Intercept)     Velocity
0.1186661507 0.0002278214

> (Cmat<- matrix(c(1,0),nrow=1,ncol=p))## the only change!

     [,1] [,2]
[1,]    1    0
```

```
> (CBhat<- as.vector(Cmat%*%betahat))

[1] 0.3991704

> (CBse<- as.vector(sqrt(Cmat %*% varBhat %*% t(Cmat))))

[1] 0.1186662

> (tmult<- qt(1-0.05/2, df=n-p))

[1] 2.073873

> CBhat + c(-1,1) * CBse * tmult

[1] 0.1530719 0.6452690
```

(b) Use the `confint` (and `lm`) function to verify your interval estimate, above. ANS:

```
> confint(lm(Distance ~ Velocity, data=bigbang.df),
+         parm=1)

                2.5 %   97.5 %
(Intercept) 0.1530719 0.645269
```

Bibliography

[Wak13] Jon Wakefield. *Bayesian and Frequentist Regression Methods*. Springer, New York, 2013.